LLNL MIDDLE EAST AND NORTH AFRICA AND FORMER SOVIET UNION RESEARCH DATABASE

Jennifer L. O'Boyle, Stanley D. Ruppert, Teresa F. Hauk, Doug Dodge, Mike Firpo
Lawrence Livermore National Laboratory

ABSTRACT:
The Lawrence Livermore National Laboratory (LLNL) Ground-Based Nuclear Explosion Monitoring (GNEM) R&D program has made significant progress populating a comprehensive Seismic Research Knowledge Base (SRKB) and deriving calibration parameters for the Middle East and North Africa (ME/NA) and Former Soviet Union (FSU) regions. The LLNL SRKB provides not only a coherent framework in which to store and organize very large volumes of collected seismic waveforms, associated event parameter information, and spatial contextual data, but also provides an efficient data processing/research environment for deriving location and discrimination correction surfaces. The SRKB is a flexible and extensible framework consisting of a relational database (RDB), Geographical Information System (GIS), and associated product/data visualization and data management tools. This SRKB framework is designed to accommodate large volumes of data (over 2 million waveforms from 20,000 events) in diverse formats from many sources (both LLNL derived research and integrated contractor products) in addition to maintaining detailed quality control and metadata. Using the SRKB framework, we are combining travel-time observations, event characterization studies, and regional tectonic models to assemble a library of ground truth information and phenomenology (e.g. travel-time and amplitude) correction surfaces required for support of the ME/NA and FSU regionalization program. Corrections and parameters distilled from the LLNL SRKB provide needed contributions to the DOE Knowledge Base (DOE KB) for the ME/NA and FSU regions and will help improve monitoring for underground nuclear testing. The LLNL research products will facilitate calibration of International Monitoring System (IMS) stations (primary and auxiliary), their surrogates (if not yet installed) and selected gamma stations necessary to complete the above tasks in the ME/NA and FSU regions. We present expanded lookup tables for critical station parameter information (including location and response) and a new integrated and reconciled event catalog dataset, including specification of preferred origin solutions and associated phase arrivals for the complete PDE, EDR, CMT, ISC bulletins and selected regional catalogs. In addition to an overview of selected datasets and individual research products, we present an overview of our visualization, integration, and organizational processes. Development of these processes and the LLNL SRKB was necessitated by both the very large amount of data and information involved (over 1.5 TB) and the varied data and research result formats utilized. Products contained and organized within the LLNL SRKB are grouped in 5 major categories:

     1) Reference contextual information and lookup tables
     2) Ground truth and regionalization data
     3) Event location products
     4) Event identification products
     5) Visualization and data manipulation processes

	Form Approved
Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **SEP 2000**	2. REPORT TYPE	3. DATES COVERED **00-00-2000 to 00-00-2000**

4. TITLE AND SUBTITLE **LLNL Middle East And North Africa And Former Soviet Union Research Database**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Lawrence Livermore National Laboratory,Livermore,CA,94550**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Proceedings of the 22nd Annual DoD/DOE Seismic Research Symposium: Planning for Verification of and Compliance with the Comprehensive Nuclear-Test-Ban Treaty (CTBT) held in New Orleans, Louisiana on September 13-15, 2000. U.S. Government or Federal Rights.

14. ABSTRACT
See Report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**10**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

OBJECTIVE

The primary objective of the Lawrence Livermore National Laboratory (LLNL) Seismic Research Knowledge Base (SRKB) is to help coordinate the LLNL Ground-Based Nuclear Explosion Monitoring (GNEM) Middle East, North Africa (ME/NA) and Former Soviet Union (FSU) regionalization program. The LLNL SRKB provides efficient access to, and organization of, thousands of seismic events and associated waveforms, while also providing the framework to store, organize, and disseminate research results for delivery into the DOE KB. Reference event libraries and ground truth datasets showing space and time clustering of natural earthquake and mine events, phase blockage maps, and event characterization parameters are compiled from the SRKB. Sufficient metadata (including measurement procedures, codes, comments and measurement errors) are stored at each step in the analysis process to allow recreation or verification of results at any stage in the processing flow. Corrections and parameters distilled from the LLNL SRKB provide needed contributions to the Department of Energy (DOE) Knowledge Base (KB) for the ME/NA and FSU regions and will enable the United States National Data Center (USNDC) to effectively verify Comprehensive Nuclear-Test-Ban Treaty (CTBT) compliance. The LLNL contributions support critical USNDC pipeline functions in detection, location, feature extraction, discrimination, and analyst review. Figure 1 outlines the LLNL GNEM research processes and integration of research into the DOE KB.

LLNL research products created using the SRKB may be grouped under two major categories: primary data products and derived products. The primary products are those developed in the process of collecting the raw materials for calibration: ground truth data, waveform data, event catalogs, phase pick information, regional station information and instrument responses. The derived products (distilled from the organized raw seismological data) are models and corrections that improve detection, location and discrimination functions. To calibrate International Monitoring System (IMS) stations (primary and auxiliary), as well as a variety of gamma stations that experience has shown to be useful, the LLNL SRKB must incorporate and organize the following categories of primary and derived measurements, data and metadata:

Contextual and Reference Datasets
1) Station Parameter Tables and Instrument Responses
2) Global and Regional Earthquake Catalogs
3) Selected Calibration Events
4) Geologic/Geophysical Datasets
5) MENA 1.1 Geophysical Model

Measurements and Raw Data
1) Event Waveform Data
2) Phase Picks
3) Travel-time and Velocity Models
4) Rayleigh and Love Surface Wave Group Velocity Measurements
5) Phase Amplitude Measurements and Magnitude Calibrations
6) Middle East Mine Atlas and Reference Waveforms

In order to efficiently organize this information within the SRKB, it was necessary to automate procedures needed to create and update database tables. This need prompted the development of custom tables to track additional information created by or related to loading data. We are also in the process of developing custom database tables (where existing defined tables are inadequate) in order to organize and store research results in a manner that will allow these results to be accessible to the research team and to track metadata related to table entries.

RESEARCH ACCOMPLISHED

Contextual Data Acquisition

Data collection for the LLNL SRKB began in 1996 and continues today. The LLNL study area surrounding the Middle East and North Africa including Europe, part of the Former Soviet Union, and Central Africa, has been termed the ME/NA and FSU regions (between latitudes 10° South and 90° North and longitudes -20° West and 85° East). Most of the archived waveforms in the LLNL SRKB are from events located within this region and occurring between 1976 and 2000. Special datasets are compiled for specific areas of interest or special events, such as the Novaya Zemlya test site and the 1998 India and Pakistan nuclear tests. The SRKB is organized in CSS3.0 (Center for Seismic Studies Version 3.0) format with USNDC and LLNL extensions.

Station Parameters

Development of travel-time corrections requires not only accurate source locations, origin times and quality-controlled waveform data, but also accurate knowledge of station locations. Seismic station information is a metadata requirement needed to support all stages of seismic waveform analysis. These metadata include such parameters as station operation dates, location and elevation, type of channels and instruments, sampling rates, and instrument responses. Our main source of this information is Incorporated Research Institute for Seismology (IRIS) "dataless" SEED files, which are provided by each of the networks affiliated with IRIS. These files contain all station parameters and response information and are updated periodically by the networks. Other station information has been obtained through Internet station books and AutoDRM systems. CSS3.0 **site** and **sitechan** table entries (listing station location, available channels, sensor orientations, operation dates, etc.) were created for almost all IRIS affiliated networks as well as many other stations with waveforms in the SRKB (Figure 2).

As networks become more expansive and install more instruments, we have encountered some confusion over seismic stations with multiple operators and duplication of channels or simpler cases where two stations were given the same code. To resolve these problems of non-uniqueness in our database tables, we have appended a network code to some of the station codes. We attempt to standardize the channel names based on the FDSN naming convention used at IRIS. However, consistent adherence to the FDSN naming convention was not possible in all cases. In the future, we intend to append a 2-character location code that is now being utilized by some IRIS networks to the channel code. This is necessary to distinguish between multiple channels of the same type operating simultaneously at a single station.

We reviewed all new station and channel information and existing database entries for completeness and coherency before updating the SRKB tables. Over 1800 station and array element table entries have been updated, but we still have not located reliable parameter data for some stations. Minimal or inconsistent information restricts the reliability of some entries. Often discrepancies arise between multiple information sources for the same station, but IRIS or the network operator is assumed to offer the most complete information in most cases. We maintain a list of sources for each station parameter in custom database tables, since each parameter may be obtained from a different source. **Network** and **affiliation** tables have been created to track the network(s) to which each station is associated.

Instrument and **sensor** tables are used to document instrument type and response for each station and channel. The IRIS "dataless" SEED files are used to generate instrument response (RESP) files for each station/network/channel/time combination delineated by calibration periods. The **sensor** table references the response file through the **instrument** table, which contains pointers to the online flatfiles. Frequency-amplitude-phase (FAP) files have been provided by some other networks and are similarly stored with database pointers. The Seismic Analysis Code (SAC) has been modified to interpret either response information type (RESP or FAP) for use in performing instrument response corrections on waveform data by using the EVALRESP software library available from IRIS.

To manage and update the large number of response files for stations maintained in the LLNL SRKB, we have automated the required processing and quality control steps. As we load information into the **instrument** and **sensor** tables, the automated system separates the RESP file into individual epochs, removes overlapping epochs and verifies that the epochs correspond to those in the station parameter tables. New RESP files are created from these results with the station and channel names modified if necessary. If the "frequency of sensitivity" is provided in the RESP file, the inverse value is inserted as the calibration period; otherwise, the "normalization frequency" is used. If neither value is found, the file is not processed. The IRIS EVALRESP program is used to calculate a transfer function to supply a displacement amplitude response.

Event Bulletins

Reference event locations and origin time information are necessary in most stages of our seismic processing and research. Bulletin information from many global, local, and regional earthquake catalogs has been incorporated into the LLNL SRKB and provides a much larger source for event selection. This combined and reconciled source facilitates comparison of event parameters provided by multiple networks with different degrees of location quality and accuracy for a wider range of magnitudes in comparison to the single global catalog used in the past. The global catalogs include: USGS Monthly (Final) Preliminary Determination of Epicenters (PDE) catalog, USGS Earthquake Data Report (EDR) catalog with phase arrival information, Bulletin of the International Seismological Centre (ISC) with phase arrival information, Harvard Centroid Moment Tensor (CMT) catalog, and the Multimax Ground Truth

dataset. We have also compiled numerous regional and local catalogs from countries in the ME/NA region, including Jordan, Israel, Morocco, and the Joint Seismic Observing Program (JSOP). A special catalog was compiled for events in the European Arctic and at Novaya Zemlya in collaboration with NORSAR. Figure 3 shows a global map of the event location coverage provided by several of the bulletins. We have established a number of collaborative agreements with countries and institutes in our study region that are yielding both local seismic catalogs and ground truth information as well as seismic waveform data.

We have developed an automated method to load various bulletins of different formats into native-format database tables, which retain many of the fields provided in the original bulletin, as well as assign a unique origin identifier (*orid*) to each event. Individual bulletins are then merged into a single CSS3.0 **origin** table retaining the author of the original bulletin. The event type and depth determination method are remapped to a common set of designators. A space-time correlation algorithm is used to reconcile events between individual catalogs. Events common to multiple bulletins will be assigned common event identifiers (*evid*) while retaining unique *orid*s. The **event** table stores the preferred origin for each event based on a ranked list of catalog preferences. Phase arrivals provided with native catalogs are loaded into the **assoc** and **arrival** tables. The arrivals can be used to check the event reconciliation.

Phase Information
LLNL researchers and analysts have made phase picks for over 1800 events to yield over 20,000 travel-time observations available to the LLNL research team for location and discrimination projects. Phase analysis is an ongoing effort, since new events and waveforms are continually being added to the SRKB. Phase information is recorded in the **arrival** and **assoc** tables, along with pick and waveform quality and other comments that are recorded in a **remark** table. Augmenting the LLNL picks, we added ~5.5 million USGS Earthquake Data Report (EDR) catalog phase pick observations to the SRKB to be used for travel-time correction studies and correction surface generation. Since the EDR picks are available only for 1990-1997, we supplemented the EDR picks with the complete Bulletin of the International Seismological Centre (ISC) phase arrival measurements from 1964-1996, which includes 28 million picks. Phase picks have also been entered into the SRKB from the Joint Seismic Observing Program (JSOP) and a reviewed subset of the ISC bulletin (Engdahl *et al.*, 1998).

Waveforms
We are collecting seismic data from IMS primary and auxiliary stations, as well as surrogate stations (for IMS stations not yet installed) and other stations needed to support calibration in the region of study. We have obtained up to 10 years of continuous data for important ME/NA and FSU stations from IRIS, Institut de Physique du Globe de Paris GEOSCOPE program, the GEO-Forschungs Zentrum/Potsdam, Germany GEOFON program and other data centers. Current data are being supplied by an LLNL joint project with the Jordan Natural Resources Authority from two seismometers deployed in Jordan. Data for particular events has been obtained from the prototype International Data Center (PIDC) and the USNDC. The Center for Monitoring Research (CMR) and NORSAR have provided waveforms for special regions/events, such as the Novaya Zemlya test site.

Our current data collection emphasis is on events recorded between 1990 and present, except for IMS surrogate stations no longer in operation or for special events such as nuclear tests. All available channels and components are requested from each station or array. Data from IRIS, GEOSCOPE, and GEOFON are provided in SEED format, waveforms from USNDC, NORSAR and CMR are organized in CSS format, and data from LLNL deployments are recorded in REFTEK format. Additionally, although the continuous data remain archived on tapes, seismic events are extracted from the continuous waveforms. This process is outlined in Figure 4. We are augmenting the open historic station data with seismic data from LLNL field deployments, such as the Jordanian stations, and collaborative research efforts established in ME/NA countries.

To generate the event waveforms for the SRKB, we have developed an automated way of extracting waveforms based on events selected from an **origin** table and adding entries for these waveforms into the appropriate database tables. Waveforms are extracted from continuous data in the native format and then converted to CSS3.0 format. Currently, we can perform waveform extraction on data in SAC, SEED, CSS, SEGY and GSE formats. Information already in the database is used to determine station parameters and event origins which are then used in determining the appropriate time window of data to extract. The extracted waveforms are entered into staging database tables and then are reconciled and loaded into the main database tables. At this time, waveforms are checked for poor or missing signals, which are then recorded in a metadata table. If waveforms for a specific event are segmented, the

segments will be merged together; if a gap exists between segments, the space will be filled with zero-value data points. If two segments overlap, a correlation test is done before the two segments are joined.

The number of waveforms in the SRKB is now over 2 million, which corresponds to 20,000 seismic events. In addition to individual event waveform segments and continuous data traces managed by the SRKB, we also maintain an archive of active and passive seismic data from various field deployments, which include IRIS/PASSCAL experiments in Tanzania, Pakistan, Caspian Sea, and Turkmenistan. We are also archiving the very long FSU Peaceful Nuclear Explosion (PNE) refraction profiles collected under contract to LLNL by the US Geological Survey (USGS).

SRKB waveforms may contain dropouts, glitches, timing errors or other problems affecting waveform analysis. We have developed an algorithm that analyzes and reports on 18 separate problems in the categories of timing errors, zero slope detection, discontinuity detection, and median filtering. This automated procedure measures waveform quality in a consistent manner and generates custom database tables which describe the types of problems encountered along with the location and frequency of these problems. These tables allow researchers to quickly identify sets of quality reviewed waveforms, thus minimizing the laborious process of previewing each trace.

Other Reference Datasets
Collaborations with Cornell University, USGS and other organizations allow us to obtain reference datasets useful as background and supporting information for research. Most of these datasets are provided as Geographic Information System (GIS) products, which allow us to integrate them with LLNL datasets. The datasets include geographic, geopolitical, mining industry, geologic and tectonic information.

Database Organization
The LLNL SRKB is a framework consisting of a relational database (RDB), Geographical Information System (GIS), and associated product/data visualization and data management tools, as outlined in Figure 5.   The SRKB is necessary for the storage and organization of very large volumes of collected seismic waveforms, associated event parameter information, and spatial contextual data, but also provides an efficient data processing/research environment for deriving location and discrimination correction surfaces.

Database Format
The LLNL SRKB is designed to be flexible and extensible in order to accommodate large volumes of data in diverse formats from many sources in addition to maintaining detailed quality control and metadata. The SRKB is comprised of ORACLE relational database software running on a SUN UNIX Server accessible from researcher workstations. Data are stored in CSS3.0 format (Center for Seismic Studies Version 3.0 database structure) with USNDC and LLNL extensions. These formats provide parameter defined tables for different elements of seismic data, such as event and station information, as well as allowing for customized tables to be developed for specific research needs or results. The CSS3.0 format offers the ability to organize data in a standard format, store metadata necessary for documenting research methods and deliver compatible data to DOE and other researchers.

Class Libraries
We have written a set of C++ classes designed primarily as components of computer programs for use by the LLNL GNEM program, which consists of about 130 classes. Of these, about 10 are service classes with no specific seismological capability, *e.g.* time class, architecture class, etc. All the remaining classes are specialized to some degree for use in seismological applications. All of the collection classes have database-aware specializations as well, since most of the programs developed with the classes interact extensively with the SRKB database. The class library has built-in support for the CSS3.0, SAC, GSE2.0, and PC-SUDS data formats.

Using the classes has proven to be an effective means of building the utility programs required to support the SRKB. In most cases, nearly all the required functionality exists within the classes, so the program only needs to collect user input and manage a few object interactions. Since the developer is shielded from the intricacies of data formats, embedded SQL, and the like, programs can usually be constructed quickly, and can be easily maintained and augmented. In addition, when improvements are made to any of the classes, all programs using those classes can benefit just by recompiling.

Data Access
Different researcher needs for data and metadata require that subsets of data be provided in a format easily accessible to many diverse types of software and analysis tools. Therefore, the SRKB access tools have been designed to utilize the power of the relational database to facilitate efficient queries and data retrieval. The Seismic Analysis Code (SAC) software used by LLNL researchers provides direct access to database table information and waveforms and uses the response files to perform instrument response corrections. PL/SQL language can be used to make database queries on contents of any of the available database tables.

For spatial queries and organization, we have adopted a Geographic Information System (GIS), which provides a framework to store and manipulate spatially-defined data. The GIS is linked with the ORACLE database to provide joint spatial and relational queries. We have integrated a large number of our research products and contextual datasets into the GIS, as shown in Figure 6. A customized framework provides an organized menu system for accessing the datasets and also includes additional functionality developed in collaboration with Sandia National Laboratories. The GIS is organized by geographic regions, seismic stations and research product types to provide easy access to a large number of products developed in various formats. Links are established for related non-spatial data, such as photographs or graphs, and metadata are provided for each dataset. Researchers can use the GIS to browse many of their research products and also to perform basic comparisons, queries, and analyses. It also provides a way to use LLNL products in conjunction with other datasets, including integrated contractor products.

CONCLUSIONS AND RECOMMENDATIONS
Corrections and parameters distilled from the LLNL SRKB provide needed contributions to the DOE Knowledge Base for the ME/NA and FSU regions and will enable the USNDC to effectively verify CTBT compliance. The LLNL portion of the DOE KB supports critical USNDC pipeline functions in detection, location, feature extraction, discrimination, and analyst review in the Middle East, North Africa and portions of the Former Soviet Union. A wide range of research products required to support the ME/NA and FSU regionalization program are being derived from waveforms, station parameters, and bulletin information contained in the LLNL SRKB. The derived products (distilled from the organized raw seismological data) include: models, calibrations and correction surfaces that improve detection, location, discrimination and event characterization functions, and capability studies to help guide our research efforts.

We have made a major effort to provide a product development environment that encourages the natural synergies among each of the separate research efforts of the LLNL GNEM researchers. The derivation of reference and ground truth datasets and location and identification products takes place in an integrated environment with changes and improvements in one area being used to facilitate development of the remaining areas. We utilize the LLNL SRKB as an integrating framework to provide the basis for synergistic development of all LLNL GNEM research. By combining travel-time observations, event characterization studies, and regional wave-propagation studies of the LLNL GNEM research team for ground truth and regional events, we have assembled a substantial library of ground truth information (origin times, locations, depths, magnitudes), event location correction surfaces, tomographic models and mine explosion statistics.

The datasets and research products contained and organized within the LLNL SRKB may be grouped under 5 major categories:
  1) Reference contextual information and lookup tables
  2) Ground truth and regionalization data
  3) Event location products
  4) Event identification products
  5) Visualization and data manipulation processes

Given the large quantity of data now managed by the SRKB, emphasis has shifted to create the efficient "production" level seismic data selection, processing and visualization tools necessary to meet programmatic and DOE KB needs. These tools, along with data browsers to allow visualization and quick access to both data and delivered research products, are under development in collaboration with Sandia National Laboratories. The tools provide such functionality as deriving location corrections, making amplitude and magnitude measurements, and developing discriminants.

REFERENCES
Engdahl, R., R. van der Hilst and R. Buland (1998). Global teleseismic earthquake relocation with improved travel-times and procedures for depth determination, *Bull. Seismol. Soc. Am.*, **88**, 722-743.
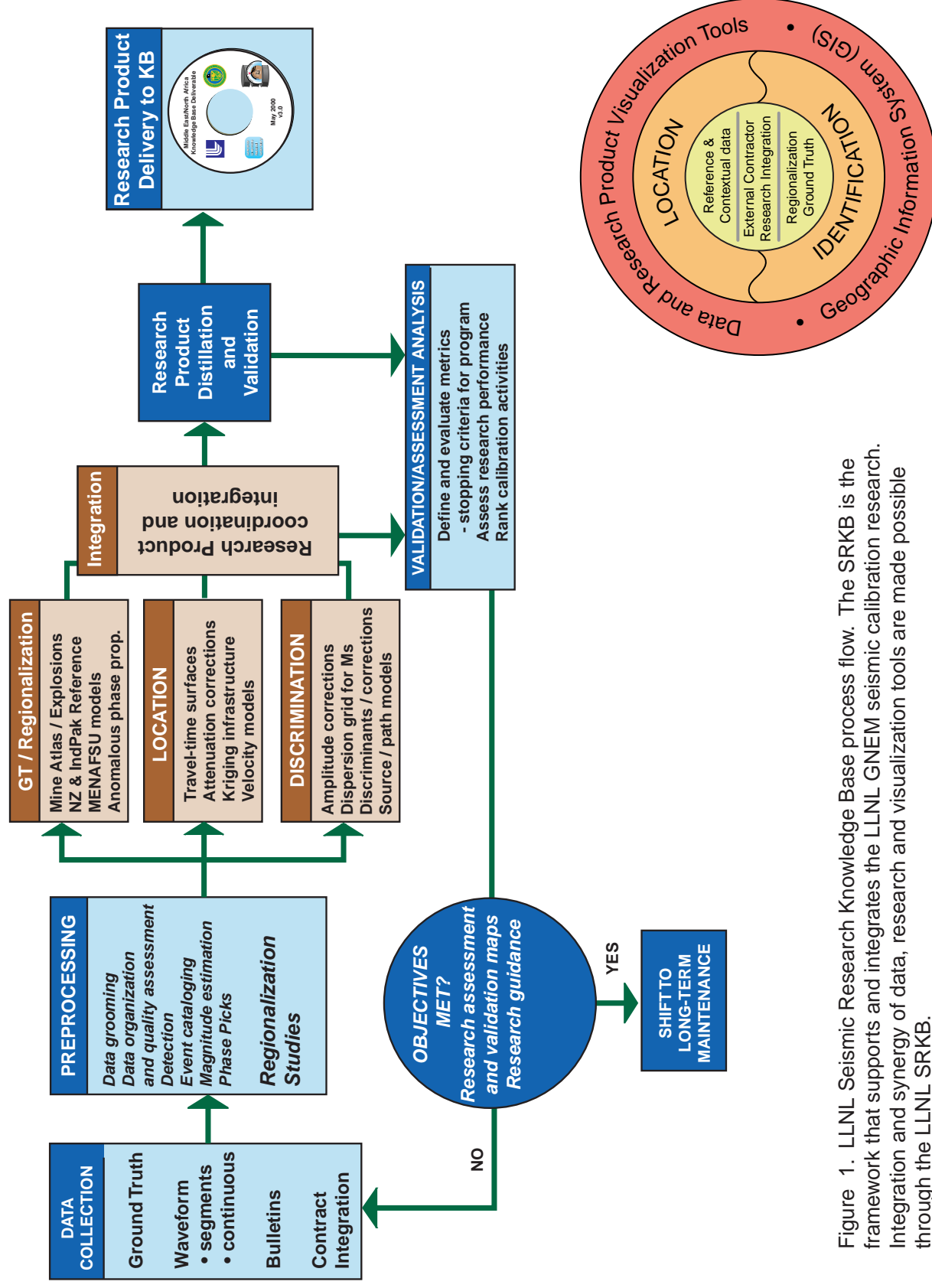
Figure 1. LLNL Seismic Research Knowledge Base process flow. The SRKB is the framework that supports and integrates the LLNL GNEM seismic calibration research. Integration and synergy of data, research and visualization tools are made possible through the LLNL SRKB.
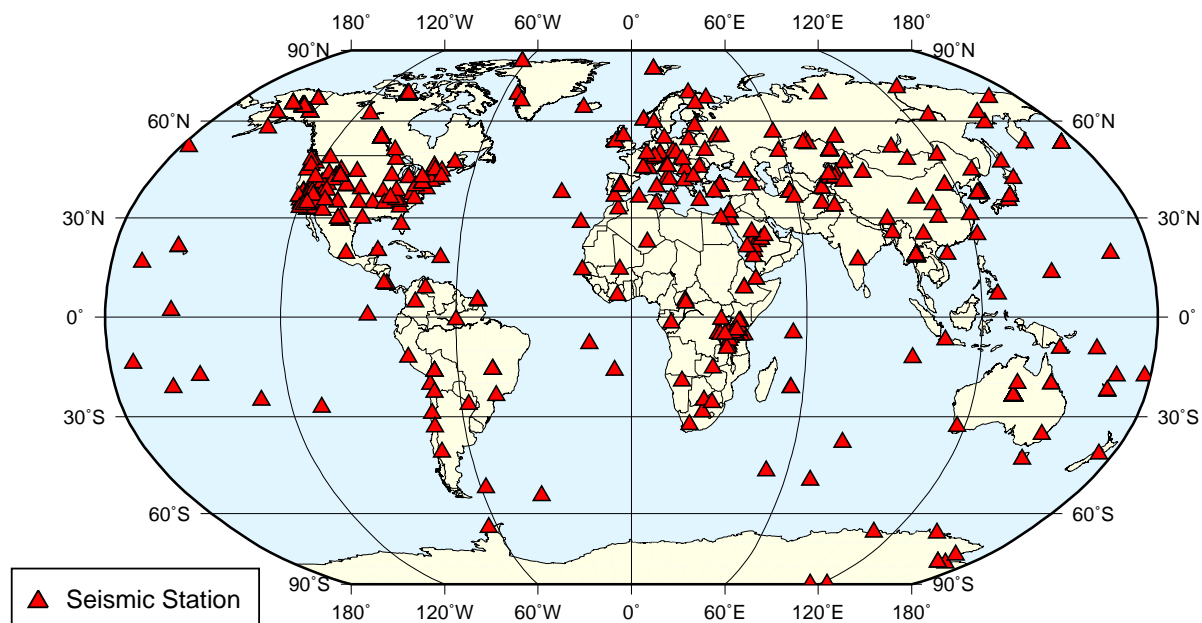
Figure 2. Distribution of seismic stations contained in the LLNL SRKB database tables. Station parameters have been reviewed and updated for most stations.
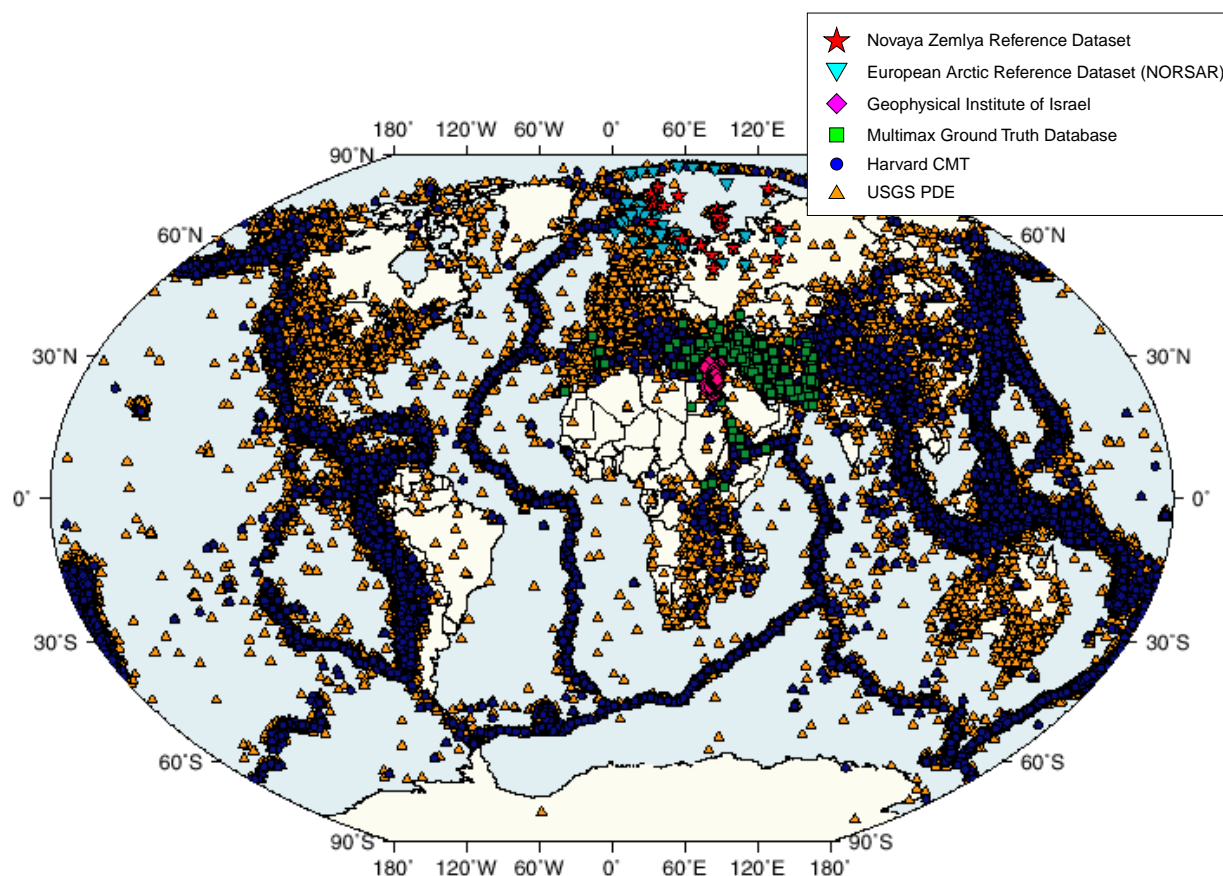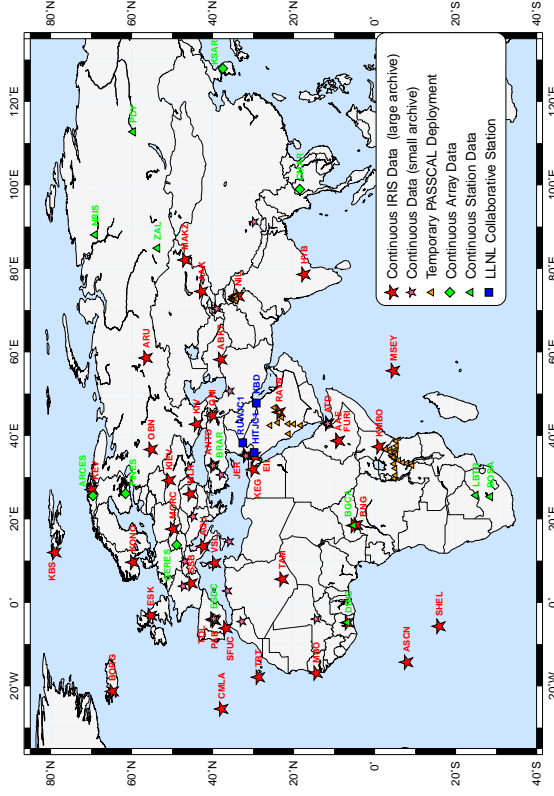

Figure 3. Events from many of the global, regional and local bulletins incorporated in the LLNL SRKB reconciled event database tables.

EVENT EXTRACTION

SEED
SAC
GSE2.0
IDC1.0
CSS
SEGY
REFTEK

Regional/Local

JSOP
Country bulletins
Aftershock bulletins

Global
PDE
EDR
ISC
CMT

Other Events
Mine blast detections
Nuclear tests
PNEs
Waveform correlation

DATABASE ENTRY

CSS3.0

WAVEFORM COLLECTION

Continuous IRIS Data (large archive)
Continuous Data (small archive)
Temporary PASSCAL Deployment
Continuous Array Data
Continuous Station Data
LLNL Collaborative Station

LLNL SRKB WAVEFORMS

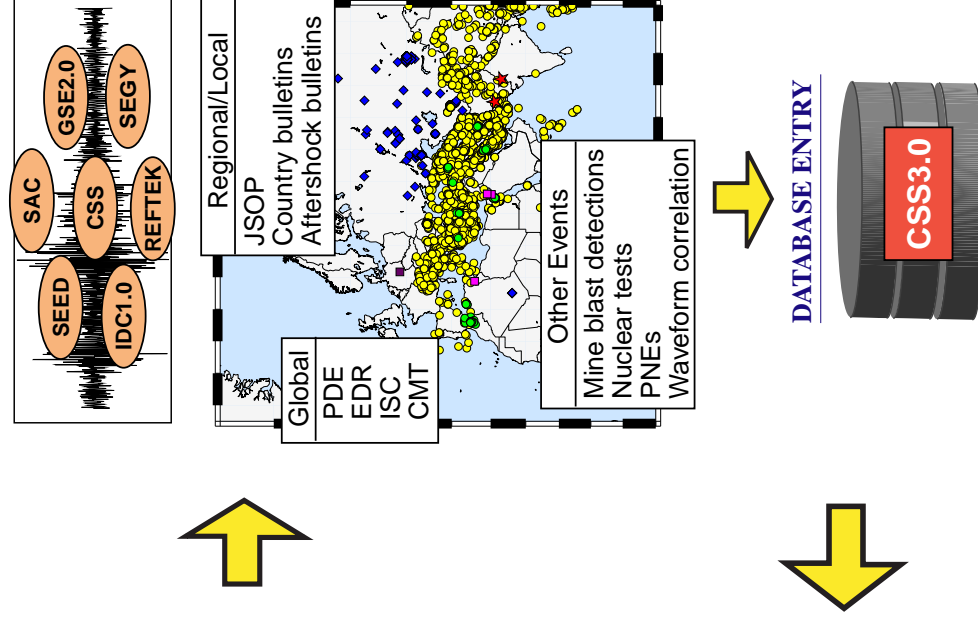Seismic Event
Seismic Station

Figure 4. Outline of the population of the SRKB with event waveforms. Continuous data are requested for a large number of stations and data are provided in various formats. Events are chosen from any number of available catalogs and the events are extracted from the continuous data. These waveforms are then converted to CSS3.0 format and loaded into ORACLE database tables.
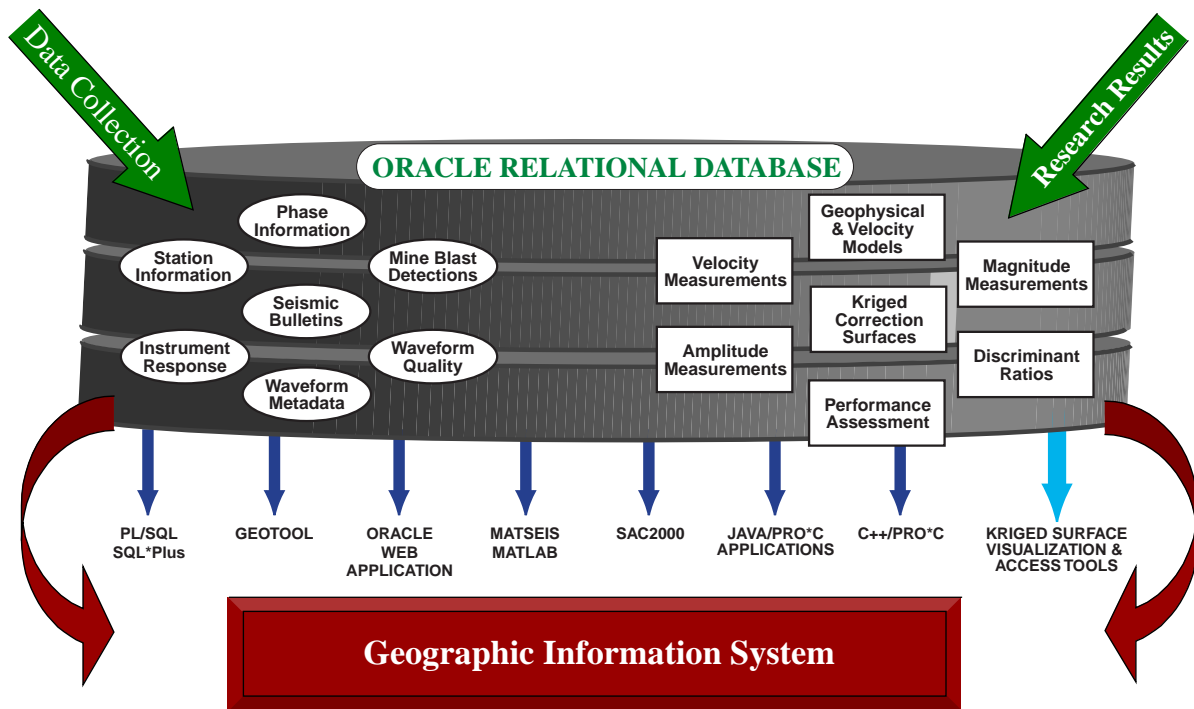
Figure 5. The LLNL SRKB integrates a relational database, a Geographic Information System (GIS) and visualization/data management processes.
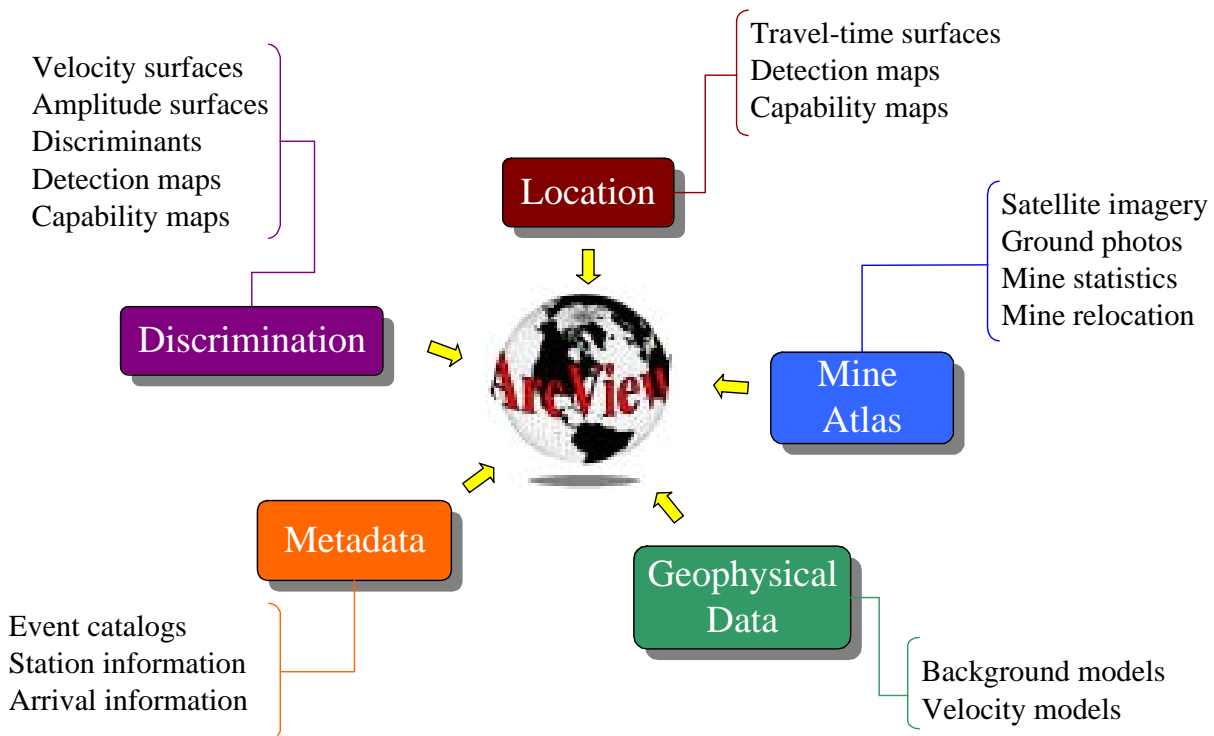

Figure 6. A large portion of raw and derived data and research products is incorporated into a GIS system. ESRI ArcView software is used to create this integrated framework for SRKB datasets.